

Mash-up of LexWiki and Web-Protégé for Collaborative Ontology Authoring

Guoqian Jiang, Harold Solbrig

Division of Biomedical Statistics & Informatics

Mayo Clinic College of Medicine

SMW 2010 Spring Conference

May 22, 2010 at MIT, Boston, MA

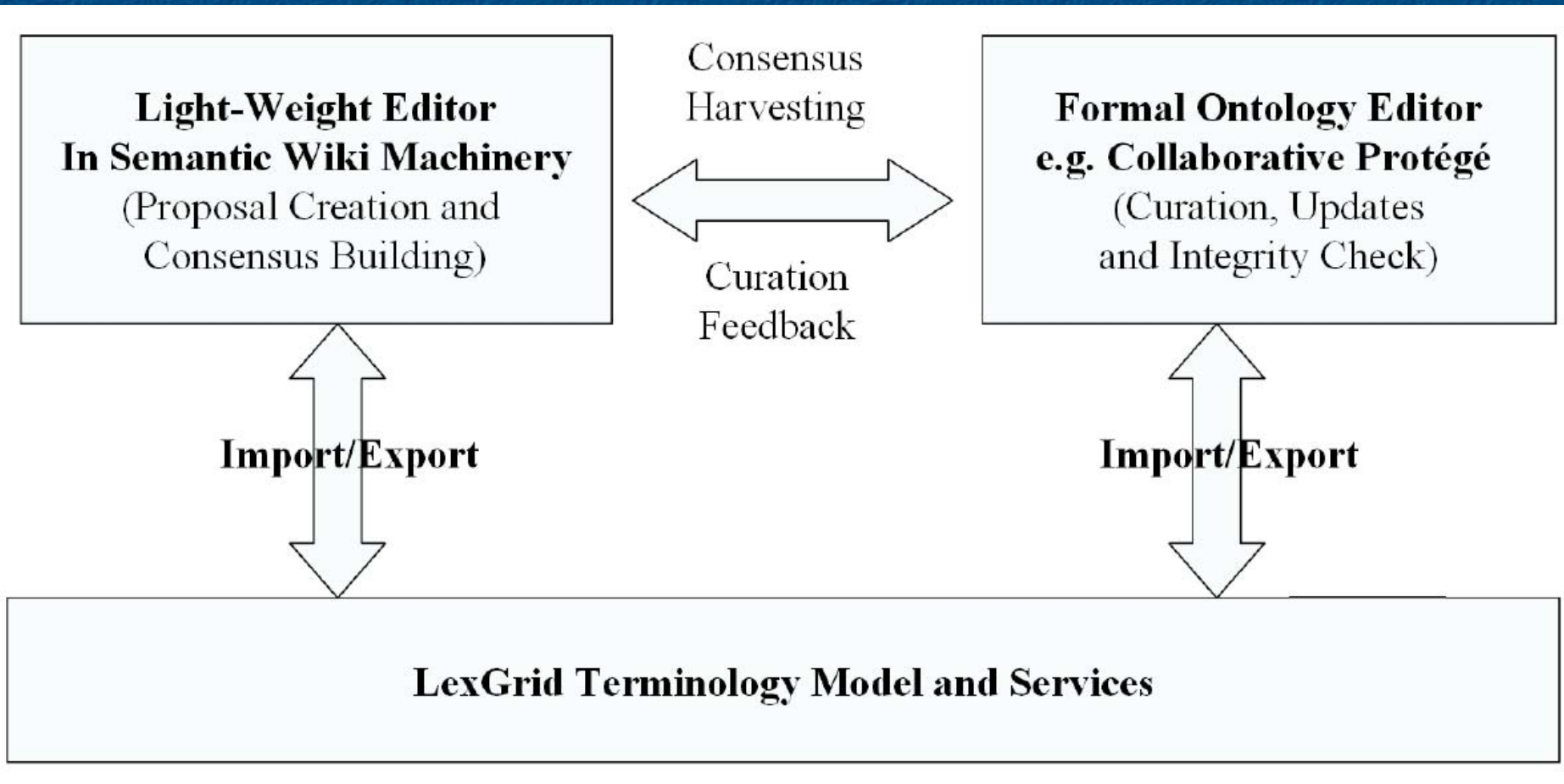
Outline

- **LexWiki**
 - LexWiki Framework
 - LexWiki Features by Use Cases
 - BiomedGT Wiki – Collaborative ontology authoring
 - LexWiki CSHARE -- Collaborative meta-data harmonization
- **Web-Protégé**
 - iCAT tool - WHO ICD11 Authoring
 - iCAT Features
- **Mash-up of LexWiki and Web-Protégé**
 - Why mash-up?
 - Prototype features
 - Challenges?

Semantic Wiki/LexWiki

- Wiki as a collaborative system – community generated content.
- Semantic wiki as an platform – support different levels of the formality continuum (Free text -> OWL).
- LexWiki - a collaborative authoring platform for large-scale biomedical terminologies
 - Implemented on Semantic MediaWiki
 - Developed by Mayo Clinic, with collaborations of NCI, WHO and Stanford University

LexWiki Framework



LexWiki Main Features

- **Ontology browsing**
- **Concept-oriented representation**
 - **Based on a formal terminology model (LexGrid model)**
- **Enhanced search functionality**
- **Proposal based workflow process**
 - **Structured proposal creation**
 - **Proposal browsing and management**
 - **Workflow package interfacing with Protégé**

LexWiki Use Cases

- **BiomedGT** – NCI Biomedical Grid Terminology
- **CTCAE** - Common Terminology Criteria for Adverse Events
- **NeuroLex** - the Neuroscience Lexicon
- **XMDR** - eXtended MetaData Registry (XMDR) Project
- **CSHARE** - CDISC Shared Health and Research Electronic Library
- **WHO ICD11** – the International Classification of Disease

BiomedGT Wiki

- An typical instance of LexWiki
- Enables groups of domain experts to collaboratively develop and maintain terminologies, including BiomedGT, subsets of BiomedGT, and other standalone terminologies.

http://biomedgt.nci.nih.gov

Main Page - BiomedGT - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://biomedgt.nci.nih.gov/wiki/index.php/Main_Page

Main Page - BiomedGT

Log in / create account

main page discussion view source history

Main Page

BiomedGT Collaborative Ontology Development Wiki

(Beta 1 Release - December 2, 2008, patch Jan 16, 2009)

About This Wiki

The BiomedGT Wiki is a new tool that enables groups of domain experts to collaboratively develop and maintain terminologies, including BiomedGT, subsets of BiomedGT, and other standalone terminologies. This wiki is being developed by the [National Cancer Institute Center for Bioinformatics](#) and the [Mayo Clinic Division of Biomedical Informatics](#) with contributions from [Apelon, Inc.](#), Northrup Grumman, and Dionne-Associates Inc.

We welcome your help in developing terminology in your specialty area and in fine tuning the use of the wiki as a collaborative terminology development tool. Several [terminology collaborations](#), such as CTCAE and NPO (see terminologies at left) are underway in this wiki.

About BiomedGT

The Biomedical Grid Terminology ([BiomedGT](#)) is an open, collaboratively developed terminology for translational research. BiomedGT builds on the strengths of the NCI Thesaurus, including concept orientation, description logic, and public accessibility. While the current terminology has been seeded with NCI Thesaurus content, it is being restructured to facilitate open content development. The goal is to evolve BiomedGT into a set of federated sub-terminologies, with content maintained by experts in the relevant research communities.

BiomedGT is registered in the [HL7 OID Registry](#) as 2.16.840.1.113883.3.26.1.3.

Becoming a Collaborator

You are welcome to browse this site and search for terminology. If you want to contribute content, you can register as a collaborator by following the instructions on the [Registration Process for Collaborators](#) page.

Subscribing to Our Listserv

You can subscribe to a listserv and receive announcements about changes to the wiki, information about support and maintenance, and other newsworthy items. To subscribe, send an e-mail message to [\[1\]](#). In the body of the message, include the following string:

```
subscribe biomedgt_wiki-1 yourname
```

Currently Available Terminologies or Terminology Subsets

- BGT - BiomedGT Grid Terminology
- Nano - Nanotechnology Domain Concepts
- CTCAE - Common Terminology Criteria for Adverse Events v4.0
- CRCH Nutrition Terms
- Biospecimen

search

Category: BGT Gene[B167]

Go Search

navigation

- Main Page
- Current events
- Recent changes
- TODO List
- Help

terminologies

- BiomedGT
- NanoTech Domain
- CRCH Nutrition Terms
- Biospecimen
- CTCAE4
- NPO
- OBI
- RadLex Draft

content search

- Query ConceptCode
- Query Preferred Name

workflow

- Create / Update Package
- Bug Reporting

toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

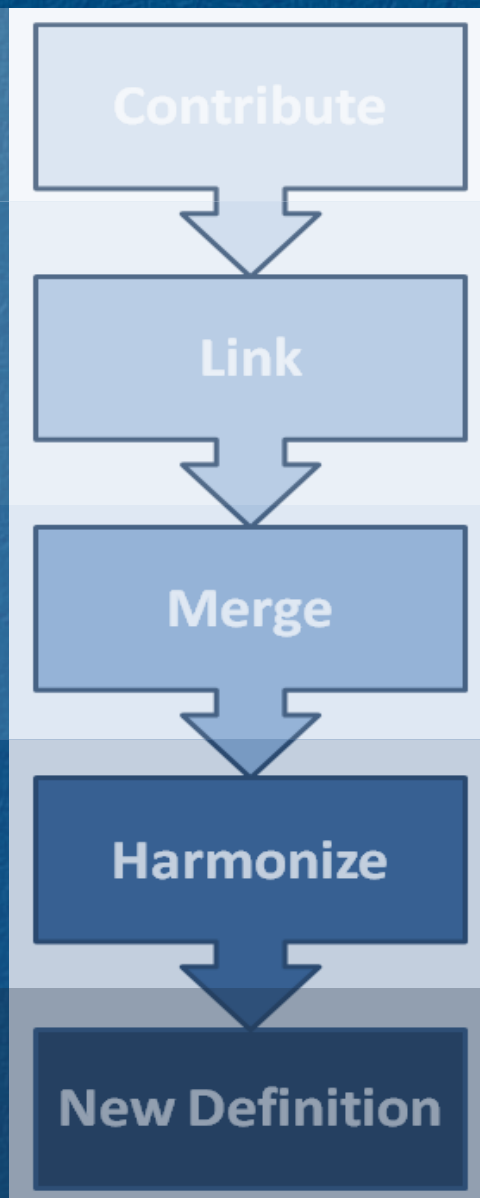
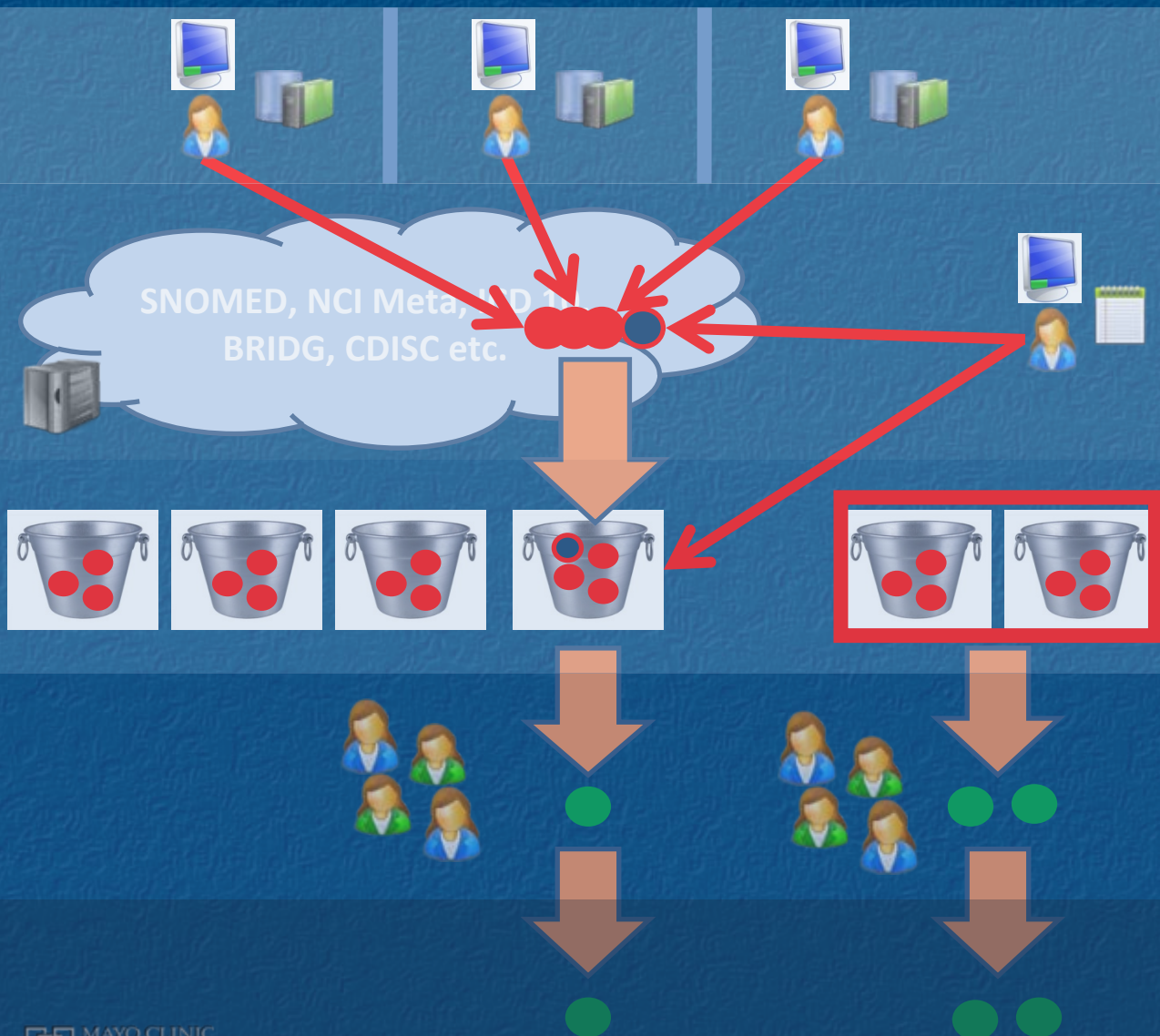
BiomedGT Demo Video (5 minutes)

- https://cabig-kc.nci.nih.gov/Vocab/uploaded_files/e/e7/Bgtwiki_part2.swf

LexWiki CSHARE

- The CDISC Shared Health and Research Electronic Library (CSHARE) wiki is a collaborative authoring environment that enables community members to harmonize data elements from multiple organizations.
- CSHARE wiki is an extension of the LexWiki platform.

CSHARE Harmonization Process



<http://informatics.mayo.edu/cshareDemo> (cshare/share)

CSHARE - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://informatics.mayo.edu/cshareDemo/index.php/Main_Page

CSHARE

CDISC Wiki

129.176.151.28 Talk for this IP Log in / create account

NICHD EDUCATION LEVEL Value Set > 11179 Enumerated Value Domain > NICHD EDUCATION LEVEL Value Set > CSHARE Value Set > Main Page >

page discussion view source history refresh

Main Page

CSHARE Demo

cDISC Shared Health And Clinical Research Electronic Library Demonstration Wiki

This is not the online CSHARE Wiki - this wiki is an approximate image of the online wiki to be used for demonstration purposes.

Note: We have uncovered some browser incompatibilities in a couple of the wiki extensions. If you don't see a search box in the upper right hand corner of this page, or if the BRIDG data element viewer doesn't have several boxes on the right hand side, you may be using an incompatible browser version. [Check here](#) for more information.

Objectives

The objective is to provide a Semantic Wiki based platform and tools, terminologies and other resources needed for the CDISC community to be able to evaluate the potential viability of a wiki-based approach for harmonizing data elements. The focus is on gaining an understanding of the recurring cost and associated benefits of using a Web 2.0 style platform for harmonization.

Presentation 12/16/2009

Content

Genzyme GSK Eli Lilly Mayo MD Anderson Common Data Elements

- Genzyme Adverse Events AE New Severity Grade
- Genzyme Adverse Events AE Severity Grade change
- Genzyme Adverse Events AE unique ID
- Genzyme Adverse Events Action Taken with Study Treatment
- Genzyme Adverse Events Causality
- Genzyme Adverse Events Check if DLT
- Genzyme Adverse Events DCODE
- Genzyme Adverse Events End Day of Adverse Event
- Genzyme Adverse Events End Hour of Adverse Event
- Genzyme Adverse Events End Minutes of Adverse Event
- Genzyme Adverse Events End Month of Adverse Event
- Genzyme Adverse Events End Year of Adverse Event
- Genzyme Adverse Events Grade Change Day
- Genzyme Adverse Events Grade Change Hour
- Genzyme Adverse Events Grade Change Minutes
- Genzyme Adverse Events Grade Change Month
- Genzyme Adverse Events Grade Change Year
- Genzyme Adverse Events New Seriousness
- Genzyme Adverse Events Other Action Taken
- Genzyme Adverse Events Other Action Taken - Hospitalization
- Genzyme Adverse Events Other Action Taken - Medication
- Genzyme Adverse Events Other Action Taken - Non Drug Therapy

Find: Next Previous Highlight all Match case

Done

Search LexEVS ☐

Slicing and Dicing (Exhibit)

page discussion edit history delete move protect watch refresh

CSHARE Lesion Measurement

TILES • TABLE

sorted by: [page name](#); then by... ☒ grouped as sorted

1. Genzyme Lesion Measurement METHOD

domain: Oncology Lesion Measurement
source: Genzyme
datatype: DT CD
concept: NCIM Measurement method(C1299991)
bridg mapping: BRIDG PerformedObservation methodCode
is mapped: true
maps from: CDE Lesion Measurement Method of Measurement, coded
value set: Genzyme METHOD6 Value Set
long name: METHOD
short name: METHOD
definition: Method used to measure the specified lesion

2. GSK Lesion Measurement Method

domain: Oncology Lesion Measurement
source: GSK
datatype: DT CD
concept: NCIM Measurement method(C1299991)
bridg mapping: BRIDG PerformedObservation methodCode
is mapped: true
maps from: CDE Lesion Measurement Method of Measurement, coded

Source

1 Genzyme
1 GSK
1 Mayo

is mapped

1 false
3 true

datatype

3 DT CD
1 DT INT
1 DT ST

concept

size(C0449453)
1 NCIM Lesion(C0221198)
3 NCIM Measurement method(C1299991)
1 NCIM Palnabla(C0522499)

bridg mapping

3 BRIDG PerformedObservation methodCode

New Definition

[page](#)[discussion](#)[edit with form](#)[edit](#)[history](#)[delete](#)[move](#)

CDE Lesion Measurement Method of Measurement, coded

[Harmonization](#)[Description](#)[Clump](#)[ConceptReference](#)

Equivalent Data Elements

Property Name	Property Value	Matchieness	Comments
Data Element From Genzyme:	Genzyme Lesion Measurement METHOD	(none)	(none)
Data Element From GSK:	GSK Lesion Measurement Method	(none)	(none)
Data Element From Eli Lilly:	EliLilly Lesion Measurement Lesion Method of Measurement	(none)	(none)
Data Element From Mayo:	Mayo Lesion Measurement Method of Evaluation	(none)	(none)
Data Element From BRIDG:	BRIDG PerformedObservation methodCode	(none)	(none)

(Curation Status: [InProgress](#))

Categories: [CSHARE WorkflowCurationStatusInProgress](#) | [CDE Data Element](#)

Integration of Halo Data Import with Exhibit (Tapping Remote Web Services)

Page

Discussion



```
{(#ws:OpenMDR_Testing1  
|term=Date  
|?result.deid  
|?result.preferred  
|?result.definition  
|?result.contextName  
|?result.objectClassName  
|?result.propertyName  
|_format=exhibit  
}}
```

Exhibit Browser

Page

Discussion

Refresh

History

Edit

?

More ▾

TABLE • TILES

2 Items

deid	preferred	definition	contextName ▾
cagrid.org_aaadda29-addb-42df-97a7-242ed0092bc4_0.1	Actual Enrollment End Date	Date on which the last study participant was enrolled	HSDB
cagrid.org_b57fe969-75b6-44e3-86dc-b53b16308da9_0.1	Actual Enrollment Start Date	Date on which the first study participant was enrolled	HSDB

CDE Generator

RDF/XML

Semantic wikitext

Tab Separated Values

Exhibit JSON

Generated HTML of this view

browsing controls here...

Date

deid

1 cagrid.org_aaadda29-addb-42df-97a7-242ed0092bc4_0.1

1 cagrid.org_b57fe969-75b6-44e3-86dc-b53b16308da9_0.1

preferred

1 Actual Enrollment End Date

1 Actual Enrollment Start Date

definition

1 Date on which the first study participant was enrolled

1 Date on which the last study participant was enrolled

contextName

2 HSDB

objectClassName

2 Study

propertyName

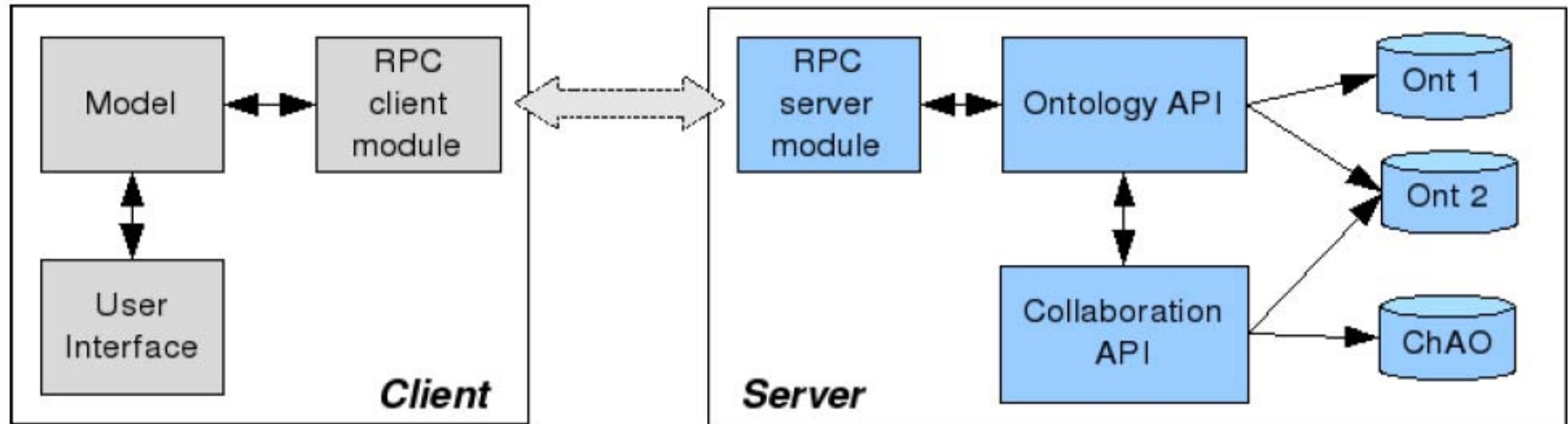
1 Enrollment End Date

1 Enrollment Start Date

Web-Protégé

- An open source light weight ontology editor for the Web that uses Protégé as its backend.
- Different users may edit the same ontology either in Web-Protégé or in a Collaborative Protégé desktop client and they will see each others changes right away.

Simplified Architecture of Web-Protégé



- **Client side**
 - Google web toolkit (GWT) for the user interface
 - Allows to write a web front end in Java and then compile into highly optimized JavaScript.
- **Server side**
 - Provides ontology-access services through the Ontology API
 - Provides support for collaboration services
 - Annotation of ontology components
 - Change tracking



My WebProtégé Collaborative Pizza

Classes Properties Individuals Notes and Discussions Other Terminologies Test Properties Manage Hierarchy

Add content to this tab Add tab

Classes

Create Delete Watch

Search: Type search string

- owl:Thing
 - Class_2
 - DomainConcept
 - Country
 - IceCream
 - Pizza
 - CheeseyPizza
 - InterestingPizza
 - MeatyPizza
 - NamedPizza
 - NonVegetarianPizza
 - RealItalianPizza
 - SpicyPizza
 - SpicyPizzaEquivalent
 - VegetarianPizza
 - VegetarianPizzaEquivalent1
 - VegetarianPizzaEquivalent2
 - PizzaBase
 - PizzaTopping
 - qweqweq
 - ValuePartition

Parents of CheeseyPizza

Pizza

Properties for CheeseyPizza

Add property value Delete property value

Property	Value	Lang
rdfs:comment	Any pizza that has at least 1 cheese topping.	en
rdfs:label	PizzaComQueijo	pt

Axioms for CheeseyPizza

Equivalent classes (Necessary and Sufficient conditions)

Pizza
hasTopping *some* CheeseTopping

Notes Tree for CheeseyPizza

New Topic

Notes for CheeseyPizza

New Topic Reply Delete Expand

<Previous Displaying page 0 of 0 pages Next>

Subject	Author	Type	Date
---------	--------	------	------

iCAT Tool – WHO ICD11 Authoring

- iCAT is the name of WHO ICD collaborative authoring tool.
- ICD11 alpha revision phase
 - iCAT is a customization of Web-Protégé developed by Stanford University.
 - It is a web based application using Google Web Toolkit (GWT) technology.
- ICD11 beta revision phase
 - New requirements
 - Public review
 - Proposal based workflow
 - Mash-up of LexWiki and Web-Protégé
 - Mayo's proposal

iCAT Main Features

- ICD content model (customized) browsing and editing
- Community collaboration including issue discussion, peer review. User can add notes and discussions to a category or a term attached to a category.
- Linkage to other terminologies (e.g. SNOMED, GO, etc.) through BioProtal ontology service
- Hierarchy management including moving in hierarchy, adding/removing parents, creating class and retiring class).
- Change history tracking based on Change and Annotation Ontology (ChAO) mechanism.
- Limited workflow support, e.g. modify access policies

Other Features

- Leveraging the strength of both GWT and collaborative Protégé, other features of the tool include:
 - very customizable
 - enforces access policies
 - many other graphical widgets available
 - can open multiple ontologies simultaneously
 - can work on the same copy of an ontology as Protégé/Collaborative Protégé
 - Plug-in architecture - easy to create your own views and extensions

ICD Categories

Create Watch Search:

- ICD Categories 1 3384
 - 01 I Certain infectious and parasitic diseases 441
 - 02 II Neoplasms 296
 - C00-C97 Malignant neoplasms 130
 - D00-D09 In situ neoplasms 36
 - D10-D36 Benign neoplasms 114
 - D37-D48 Neoplasms of uncertain or unknown behaviour 17
 - 03 III Diseases of the blood and blood-forming organs and certain disorders involving the
 - 04 IV Endocrine, nutritional and metabolic diseases 57
 - 05 V Mental and behavioural disorders 87
 - 06 VI Diseases of the nervous system 484
 - 07 VII Diseases of the eye and adnexa 2 353
 - 08 VIII Diseases of the ear and mastoid process 2
 - 09 IX Diseases of the circulatory system 21
 - 10 X Diseases of the respiratory system 25
 - 11 XI Diseases of the digestive system 331
 - 12 XII Diseases of the skin 3 1605
 - 13 XIII Diseases of the musculoskeletal system and connective tissue 128
 - 14 XIV Diseases of the genitourinary system 160
 - 15 XV Pregnancy, childbirth and the puerperium 40
 - 16 XVI Certain conditions originating in the perinatal period 14
 - 17 XVII Congenital malformations, deformations and chromosomal abnormalities 42
 - 18 XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere c
 - 19 XIX Injury, poisoning and certain other consequences of external causes 5

Parents of C00-C97 Malignant neoplasms

02 II Neoplasms

Details for C00-C97 Malignant neoplasms

Title & DefinitionType & UseTermsClinical DescriptionManifestation PropertiesCausal PropertiesTemporal PropertiesSeverity PropertiesFunctional PropertiesSpecific Condition PropertiesTreatmentSNOMED ReferencesICD 10 Notes and Hints

ICD 10 Code ?C00-C97✕💬

Sorting labelC00-C97✕💬

ICD Title ?Malignant neoplasms✕💬

Definition ?

Text

+ Add new value

External Definitions ?

+ Add new value

Definition	CUI	Ontology ID
A tumor composed of atypical neoplastic, often pleomorphic cells that invade other tissues. Malignant neoplasms often metastasize to distant anatomic sites and may recur after excision. The most common malignant neoplasms are carcinomas (adenocarcinomas or squamous cell carcinomas), Hodgkin and non-Hodgkin lymphomas, leukemias, melanomas, and sarcomas.		UMLS/NCI2007_05 E ✕💬
new abnormal tissue that grows by excessive cellular division and proliferation more rapidly than normal and continues to grow after the stimuli that initiated the new growth cease; tumors perform no useful body function and may be benign or malignant; benign neoplasms are a noncancerous growth that does not invade nearby tissue or spread to other parts of the body; malignant neoplasms or cancer		UMLS/CSP2006 ✕💬

Editorial Status Definition

Select a value ✕💬

Why Mash-up of LexWiki and Web-Protege?

- Scalability issue
 - In iCAT/Web-Protégé, multiple users work on the same copy of an evolving ICD category.
 - For a small community
 - For the alpha process, the user community is relatively small.
 - It works fine.
 - But, for a large community
 - In the beta phase, the ICD is supposed to be reviewed publicly and the number of user community will be huge.
 - It would not scale.

Benefits through Mash-up

- Synergize the strengths of both platforms
 - LexWiki
 - Collaborative features
 - Structured proposal creation
 - Scalability for large-scale ontology
 - Flexibility for configuration
 - Leveraging semantic web technology
 - Web-Protégé
 - Hierarchy navigation and management
 - Formal ontology model – OWL model
 - Potential for consistency checking and reasoning

Other Benefits

- **Keep the user interface consistent**
- **Sophisticated user roles**
 - **Content Experts**
 - **Classification Experts**
- **Incremental updates**
- **Cost effective**

Prototype Features

- Browse the contents of each category
- Publish the contents/updated contents of each category
- Create structured proposals for an individual category
- Harvest the proposals created in wiki platform into the iCAT environment

Prototype Demo

Challenges

- **Content synchronization**
 - OWL model vs LexWiki/LexGrid model
 - Label mangling
- **Hierarchy proposal creation**
 - Tree structure curation
- **Workflow process synchronization**
 - User account management
 - Proposal harvesting mechanism
 - Proposal review
 - Decision making to accept/ reject proposal
 - Update contents based on proposals accepted

Summary and Future works

- **Summary**

- We consider that the proposed framework is feasible and can be useful to tackle the scalability issue the terminology authoring community is facing.

- **Future works**

- Deal with those challenging issues
- Generalize the framework
 - LexWiki
 - Web-Protégé

References & Links

- LexWiki: <https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/LexWiki>.
- CSHARE Demo: [http://informatics.mayo.edu/cshareDemo/\(cshare/share\)](http://informatics.mayo.edu/cshareDemo/(cshare/share))
- Web-Protégé: <http://webprotege.stanford.edu>.
- iCAT Demo: <http://icatdemo.stanford.edu/>.
- Any questions about the presentation, email to: Jiang.guoqian@mayo.edu

Questions ?

